

## Filling in the Genomic Gaps in the Tree of Life at the Joint Genome Institute

Eileen M. Dalin<sup>1</sup>, Jan-Fang Cheng<sup>1</sup>, Susan M. Lucas<sup>2</sup>, Jonathan A. Eisen<sup>1</sup>, Hans-Peter Klenk<sup>3</sup>, and the JGI Production Sequencing Department.

<sup>1</sup>Lawrence Berkeley National Laboratory, <sup>2</sup>Lawrence Livermore National Laboratory, <sup>3</sup>DSMZ - German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7 B 38124 Braunschweig Germany.

US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA

The US Department of Energy's Joint Genome Institute (JGI) is a user facility for high-throughput sequencing that has sequenced a large number of the nearly 1000 microbial genomes completed or nearing completion. This sequence has enabled us to understand the roles of microorganisms in important processes such as bioremediation, energy production, global nutrient lifecycles, pathogenesis, and the diversity of life. However, the currently available genome sequences are heavily biased based upon their phylogenetic distribution when compared to the microbial genome diversity known to exist. To overcome this bias, the JGI has initiated a project known as the Genomic Encyclopedia of Bacteria and Archaea (GEBA). The long-term goal of the GEBA project would be to generate reference genomes for every major and minor group of bacteria and archaea. To test the feasibility of the project, JGI has launched a pilot project in collaboration with DSMZ to sequence 100 bacterial and archaeal genomes based on their phylogenetic positions in the tree of life. The pilot project could expand to represent 5,000 genomes. Selection of the microbes was based upon analysis of the rRNA tree of life as well as consultation with a scientific advisory board. Partaking in a project so large will present unique obstacles. The JGI will use this pilot study to understand and establish new processes to deal with the potential number of samples this long-term project might include. GEBA's pilot project progress, challenges, and developments will be presented.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344 and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

UCRL-ABS-236472